

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-268699

(43)Date of publication of application : 20.09.2002

(51)Int.Cl.

G10L 21/04  
A63H 11/00  
G10L 13/08  
G10L 13/00  
G10L 13/06

(21)Application number : 2001-066376

(22)Date of filing : 09.03.2001

(71)Applicant : SONY CORP

(72)Inventor : YAMAZAKI NOBUHIDE  
KOBAYASHI KENICHIRO  
ASANO KOJI  
KARIYA SHINICHI  
FUJITA YAEKO

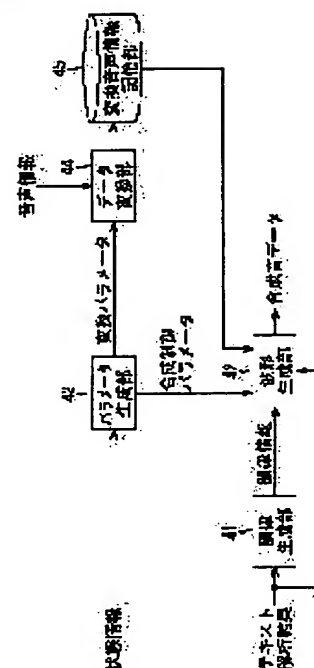
## (54) DEVICE AND METHOD FOR VOICE SYNTHESIS, PROGRAM, AND RECORDING MEDIUM

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To obtain a synthesized voice which is rich in feeling by generating the synthesized voice whose voice quality is changed according to the state of feelings.

**SOLUTION:** A parameter generation part 43 generates conversion parameters and synthesis control parameters according to state information showing the state of feelings of a pet robot. A data conversion part 44 converts frequency characteristics of phoneme piece data as voice information. A waveform generation part 42 obtains necessary phoneme piece data according to vocal sound information included in a text analysis result, processes and connects the phoneme piece data according to the vocal sound information and synthetic control parameters to generate synthesized voice data of the corresponding vocal sound and voice quality.

図6



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

# (12) 公開特許公報 (A)

(11) 特許出願公開番号  
**特開2002-268699**  
(P 2 0 0 2 - 2 6 8 6 9 9 A)  
(43) 公開日 平成14年 9月20日 (2002. 9. 20)

(51) Int. Cl. <sup>7</sup>	識別記号	F I	テーマコード (参考)
G10L 21/04		A63H 11/00	Z 2C150
A63H 11/00		G10L 3/02	A 5D045
G10L 13/08		3/00	H
13/00			Q
13/06		5/04	F
審査請求 未請求 請求項の数10 O L (全18頁)			

(21) 出願番号 特願2001-66376 (P 2001-66376)

(22) 出願日 平成13年 3月 9日 (2001. 3. 9)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川 6丁目 7番35号

(72) 発明者 山崎 信英

東京都品川区北品川 6丁目 7番35号 ソニ  
ー株式会社内

(72) 発明者 小林 賢一郎

東京都品川区北品川 6丁目 7番35号 ソニ  
ー株式会社内

(74) 代理人 100082131

弁理士 稲本 義雄

最終頁に続く

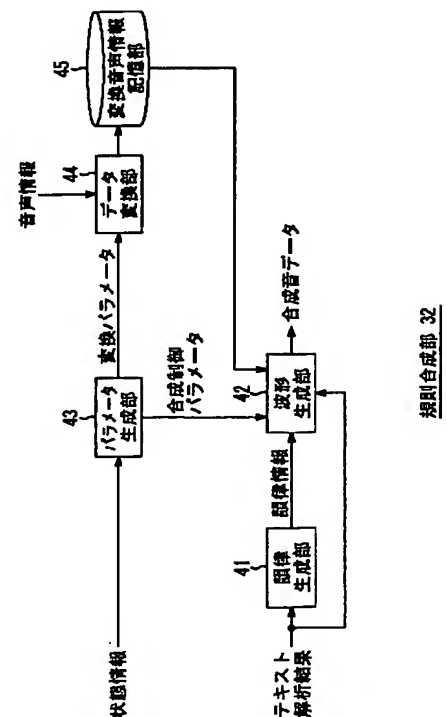
(54) 【発明の名称】 音声合成装置及び音声合成方法、並びにプログラムおよび記録媒体

## (57) 【要約】

【課題】 感情の状態に応じて音質を変えた合成音を生成することにより、感情豊かな合成音を得る。

【解決手段】 パラメータ生成部 43は、ペットロボットの感情の状態を表す状態情報に基づいて、変換パラメータと合成制御パラメータを生成する。データ変換部 44は、音声情報としての音素片データの周波数特性を変換する。波形生成部 42は、テキスト解析結果に含まれる音韻情報に基づいて、必要な音素片データを得て、その音素片データを、韻律情報と合成制御パラメータに基づいて加工しながら接続し、対応する韻律と音質の合成音データを生成する。

図6



## 【特許請求の範囲】

【請求項 1】 所定の情報を用いて音声合成を行う音声合成装置であって、  
前記所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成手段と、  
前記音質影響情報を用いて、音質を制御した前記合成音を生成する音声合成手段とを備えることを特徴とする音声合成装置。

【請求項 2】 前記音質影響情報生成手段は、  
前記感情の状態に基づいて、前記合成音を構成する波形データの特性を変更するように、前記音質影響情報を変換する変換パラメータを生成する変換パラメータ生成手段と、  
前記変換パラメータに基づいて、前記音質影響情報を変換する音質影響情報変換手段とを有することを特徴とする請求項 1 に記載の音声合成装置。

【請求項 3】 前記音質影響情報は、前記合成音を生成するのに接続される所定単位の波形データであることを特徴とする請求項 2 に記載の音声合成装置。

【請求項 4】 前記音質影響情報は、前記所定単位の波形データから抽出された特徴パラメータであることを特徴とする請求項 2 に記載の音声合成装置。

【請求項 5】 前記音声合成手段は、規則音声合成を行い、  
前記音質影響情報は、前記規則音声合成を制御するための合成制御パラメータであることを特徴とする請求項 1 に記載の音声合成装置。

【請求項 6】 前記合成制御パラメータは、音量バランス、音源の振幅ゆらぎの大きさ、または音源の周波数を制御するものであることを特徴とする請求項 5 に記載の音声合成装置。

【請求項 7】 前記音声合成手段は、周波数特性または音量バランスを制御した前記合成音を生成することを特徴とする請求項 1 に記載の音声合成装置。

【請求項 8】 所定の情報を用いて音声合成を行う音声合成方法であって、  
前記所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成ステップと、  
前記音質影響情報を用いて、音質を制御した前記合成音を生成する音声合成ステップとを備えることを特徴とする音声合成方法。

【請求項 9】 所定の情報を用いて音声合成を行う音声合成処理を、コンピュータに行わせるプログラムであって、  
前記所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成ステップと、  
前記音質影響情報を用いて、音質を制御した前記合成音

を生成する音声合成ステップとを備えることを特徴とするプログラム。

【請求項 10】 所定の情報を用いて音声合成を行う音声合成処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、  
前記所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成ステップと、  
前記音質影響情報を用いて、音質を制御した前記合成音を生成する音声合成ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関し、特に、例えば、感情豊かな合成音を生成することができるようにする音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関する。

## 【0002】

【従来の技術】従来の音声合成装置においては、テキストや発音記号を与えることによって、対応する合成音が生成される。

## 【0003】

【発明が解決しようとする課題】ところで、最近、例えば、ペット型のペットロボット等として、音声合成装置を搭載し、ユーザに話しかけるものが提案されている。

【0004】さらに、ペットロボットとしては、感情の状態を表す感情モデルを取り入れ、その感情モデルが表す感情の状態に応じて、ユーザの命令に従ったり、従わなかったりするものも提案されている。

【0005】従って、感情モデルに応じて、例えば、合成音の音質を変化させることができれば、感情に応じた音質の合成音が出力され、ペットロボットのエンタテイメント性を向上させることができると考えられる。

【0006】本発明は、このような状況に鑑みてなされたものであり、感情の状態に応じて音質を変えた合成音を生成することにより、感情豊かな合成音を得ることができるようにするものである。

## 【0007】

【課題を解決するための手段】本発明の音声合成装置は、所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成手段と、音質影響情報を用いて、音質を制御した合成音を生成する音声合成手段とを備えることを特徴とする。

【0008】本発明の音声合成方法は、所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成ステップと、音質影響情報を用いて、音質を制御した合成音を生成する音声合成ステップ

とを備えることを特徴とする。

【0009】本発明のプログラムは、所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成ステップと、音質影響情報を用いて、音質を制御した合成音を生成する音声合成ステップとを備えることを特徴とする。

【0010】本発明の記録媒体は、所定の情報のうち、合成音の音質に影響する音質影響情報を、外部から供給される、感情の状態を表す状態情報に基づいて生成する音質影響情報生成ステップと、音質影響情報を用いて、音質を制御した合成音を生成する音声合成ステップとを備えるプログラムが記録されていることを特徴とする。

【0011】本発明の音声合成装置および音声合成方法、並びにプログラムにおいては、所定の情報のうち、合成音の音質に影響する音質影響情報が、外部から供給される、感情の状態を表す状態情報に基づいて生成され、その音質影響情報を用いて、音質を制御した合成音が生成される。

【0012】

【発明の実施の形態】図1は、本発明を適用したロボットの一実施の形態の外観構成例を示しており、図2は、その電氣的構成例を示している。

【0013】本実施の形態では、ロボットは、例えば、犬等の四つ足の動物の形状のものとなっており、胴体部ユニット2の前後左右に、それぞれ脚部ユニット3A、3B、3C、3Dが連結されるとともに、胴体部ユニット2の前端部と後端部に、それぞれ頭部ユニット4と尻尾部ユニット5が連結されることにより構成されている。

【0014】尻尾部ユニット5は、胴体部ユニット2の上面に設けられたベース部5Bから、2自由度をもって湾曲または揺動自在に引き出されている。

【0015】胴体部ユニット2には、ロボット全体の制御を行うコントローラ10、ロボットの動力源となるバッテリー11、並びにバッテリーセンサ12および熱センサ13からなる内部センサ部14などが収納されている。

【0016】頭部ユニット4には、「耳」に相当するマイク（マイクロフォン）15、「目」に相当するCCD（Charge Coupled Device）カメラ16、触覚に相当するタッチセンサ17、「口」に相当するスピーカ18などが、それぞれ所定位置に配設されている。また、頭部ユニット4には、口の下顎に相当する下顎部4Aが1自由度をもって可動に取り付けられており、この下顎部4Aが動くことにより、ロボットの口の開閉動作が実現されるようになっている。

【0017】脚部ユニット3A乃至3Dそれぞれの関節部分や、脚部ユニット3A乃至3Dそれぞれと胴体部ユニット2の連結部分、頭部ユニット4と胴体部ユニット2の連結部分、頭部ユニット4と下顎部4Aの連結部

分、並びに尻尾部ユニット5と胴体部ユニット2の連結部分などには、図2に示すように、それぞれアクチュエータ3AA<sub>1</sub>乃至3AA<sub>x</sub>、3BA<sub>1</sub>乃至3BA<sub>x</sub>、3CA<sub>1</sub>乃至3CA<sub>x</sub>、3DA<sub>1</sub>乃至3DA<sub>x</sub>、4A<sub>1</sub>乃至4A<sub>L</sub>、5A<sub>1</sub>および5A<sub>2</sub>が配設されている。

【0018】頭部ユニット4におけるマイク15は、ユーザからの発話を含む周囲の音声（音）を集音し、得られた音声信号を、コントローラ10に送出する。CCDカメラ16は、周囲の状況を撮像し、得られた画像信号を、コントローラ10に送出する。

【0019】タッチセンサ17は、例えば、頭部ユニット4の上部に設けられており、ユーザからの「なでる」や「たたく」といった物理的な働きかけにより受けた圧力を検出し、その検出結果を圧力検出信号としてコントローラ10に送出する。

【0020】胴体部ユニット2におけるバッテリーセンサ12は、バッテリー11の残量を検出し、その検出結果を、バッテリー残量検出信号としてコントローラ10に送出する。熱センサ13は、ロボット内部の熱を検出し、その検出結果を、熱検出信号としてコントローラ10に送出する。

【0021】コントローラ10は、CPU(Central Processing Unit)10Aやメモリ10B等を内蔵しており、CPU10Aにおいて、メモリ10Bに記憶された制御プログラムが実行されることにより、各種の処理を行う。

【0022】即ち、コントローラ10は、マイク15や、CCDカメラ16、タッチセンサ17、バッテリーセンサ12、熱センサ13から与えられる音声信号、画像信号、圧力検出信号、バッテリー残量検出信号、熱検出信号に基づいて、周囲の状況や、ユーザからの指令、ユーザからの働きかけなどの有無を判断する。

【0023】さらに、コントローラ10は、この判断結果等に基づいて、続く行動を決定し、その決定結果に基づいて、アクチュエータ3AA<sub>1</sub>乃至3AA<sub>x</sub>、3BA<sub>1</sub>乃至3BA<sub>x</sub>、3CA<sub>1</sub>乃至3CA<sub>x</sub>、3DA<sub>1</sub>乃至3DA<sub>x</sub>、4A<sub>1</sub>乃至4A<sub>L</sub>、5A<sub>1</sub>、5A<sub>2</sub>のうちの必要なものを駆動させる。これにより、頭部ユニット4を上下左右に振らせたり、下顎部4Aを開閉させる。さらには、尻尾部ユニット5を動かしたり、各脚部ユニット3A乃至3Dを駆動して、ロボットを歩行させるなどの行動を行わせる。

【0024】また、コントローラ10は、必要に応じて、合成音を生成し、スピーカ18に供給して出力させたり、ロボットの「目」の位置に設けられた図示しないLED（Light Emitting Diode）を点灯、消灯または点滅させる。

【0025】以上のようにして、ロボットは、周囲の状況等に基づいて自律的に行動をとるようになっている。

【0026】次に、図3は、図2のコントローラ10の

機能的構成例を示している。なお、図 3 に示す機能的構成は、CPU 10A が、メモリ 10B に記憶された制御プログラムを実行することで実現されるようになっている。

【0027】コントローラ 10 は、特定の外部状態を認識するセンサ入力処理部 50、センサ入力処理部 50 の認識結果を累積して、感情や、本能、成長の状態を表現するモデル記憶部 51、センサ入力処理部 50 の認識結果等に基づいて、続く行動を決定する行動決定機構部 52、行動決定機構部 52 の決定結果に基づいて、実際にロボットに行動を起こさせる姿勢遷移機構部 53、各アクチュエータ 3A<sub>1</sub>乃至 5A<sub>1</sub>および 5A<sub>2</sub>を駆動制御する制御機構部 54、並びに合成音を生成する音声合成部 55 から構成されている。

【0028】センサ入力処理部 50 は、マイク 15 や、CCD カメラ 16、タッチセンサ 17 等から与えられる音声信号、画像信号、圧力検出信号等に基づいて、特定の外部状態や、ユーザからの特定の働きかけ、ユーザからの指示等を認識し、その認識結果を表す状態認識情報を、モデル記憶部 51 および行動決定機構部 52 に通知する。

【0029】即ち、センサ入力処理部 50 は、音声認識部 50A を有しており、音声認識部 50A は、マイク 15 から与えられる音声信号について音声認識を行う。そして、音声認識部 50A は、その音声認識結果としての、例えば、「歩け」、「伏せ」、「ボールを追いかける」等の指令その他を、状態認識情報として、モデル記憶部 51 および行動決定機構部 52 に通知する。

【0030】また、センサ入力処理部 50 は、画像認識部 50B を有しており、画像認識部 50B は、CCD カメラ 16 から与えられる画像信号を用いて、画像認識処理を行う。そして、画像認識部 50B は、その処理の結果、例えば、「赤い丸いもの」や、「地面に対して垂直なかつ所定高さ以上の平面」等を検出したときには、「ボールがある」や、「壁がある」等の画像認識結果を、状態認識情報として、モデル記憶部 51 および行動決定機構部 52 に通知する。

【0031】さらに、センサ入力処理部 50 は、圧力処理部 50C を有しており、圧力処理部 50C は、タッチセンサ 17 から与えられる圧力検出信号を処理する。そして、圧力処理部 50C は、その処理の結果、所定の閾値以上で、かつ短時間の圧力を検出したときには、「たたかれた（しかられた）」と認識し、所定の閾値未満で、かつ長時間の圧力を検出したときには、「なでられた（ほめられた）」と認識して、その認識結果を、状態認識情報として、モデル記憶部 51 および行動決定機構部 52 に通知する。

【0032】モデル記憶部 51 は、ロボットの感情、本能、成長の状態を表現する感情モデル、本能モデル、成長モデルをそれぞれ記憶、管理している。

【0033】ここで、感情モデルは、例えば、「うれしさ」、「悲しさ」、「怒り」、「楽しさ」等の感情の状態（度合い）を、所定の範囲（例えば、-1.0 乃至 1.0 等）の値によってそれぞれ表し、センサ入力処理部 50 からの状態認識情報や時間経過等に基づいて、その値を変化させる。本能モデルは、例えば、「食欲」、「睡眠欲」、「運動欲」等の本能による欲求の状態（度合い）を、所定の範囲の値によってそれぞれ表し、センサ入力処理部 50 からの状態認識情報や時間経過等に基づいて、その値を変化させる。成長モデルは、例えば、「幼年期」、「青年期」、「熟年期」、「老年期」等の成長の状態（度合い）を、所定の範囲の値によってそれぞれ表し、センサ入力処理部 50 からの状態認識情報や時間経過等に基づいて、その値を変化させる。

【0034】モデル記憶部 51 は、上述のようにして感情モデル、本能モデル、成長モデルの値で表される感情、本能、成長の状態を、状態情報として、行動決定機構部 52 に送出する。

【0035】なお、モデル記憶部 51 には、センサ入力処理部 50 から状態認識情報が供給される他、行動決定機構部 52 から、ロボットの現在または過去の行動、具体的には、例えば、「長時間歩いた」などの行動の内容を示す行動情報が供給されるようになっており、モデル記憶部 51 は、同一の状態認識情報が与えられても、行動情報が示すロボットの行動に応じて、異なる状態情報を生成するようになっている。

【0036】即ち、例えば、ロボットが、ユーザに挨拶をし、ユーザに頭を撫でられた場合には、ユーザに挨拶をしたという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部 51 に与えられ、この場合、モデル記憶部 51 では、「うれしさ」を表す感情モデルの値が増加される。

【0037】一方、ロボットが、何らかの仕事を実行中に頭を撫でられた場合には、仕事を実行中であるという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部 51 に与えられ、この場合、モデル記憶部 51 では、「うれしさ」を表す感情モデルの値は変化されない。

【0038】このように、モデル記憶部 51 は、状態認識情報だけでなく、現在または過去のロボットの行動を示す行動情報も参照しながら、感情モデルの値を設定する。これにより、例えば、何らかのタスクを実行中に、ユーザが、いたずらするつもりで頭を撫でたときに、「うれしさ」を表す感情モデルの値を増加させるような、不自然な感情の変化が生じることを回避することができる。

【0039】なお、モデル記憶部 51 は、本能モデルおよび成長モデルについても、感情モデルにおける場合と同様に、状態認識情報および行動情報の両方に基づいて、その値を増減させるようになっている。また、モデ

10

20

30

40

50

ル記憶部 51 は、感情モデル、本能モデル、成長モデルそれぞれの値を、他のモデルの値にも基づいて増減させるようになっている。

【0040】行動決定機構部 52 は、センサ入力処理部 50 からの状態認識情報や、モデル記憶部 51 からの状態情報、時間経過等に基づいて、次の行動を決定し、決定された行動の内容を、行動指令情報として、姿勢遷移機構部 53 に送出する。

【0041】即ち、行動決定機構部 52 は、ロボットがとり得る行動をステート（状態）(state)に対応させた有限オートマトンを、ロボットの行動を規定する行動モデルとして管理しており、この行動モデルとしての有限オートマトンにおけるステートを、センサ入力処理部 50 からの状態認識情報や、モデル記憶部 51 における感情モデル、本能モデル、または成長モデルの値、時間経過等に基づいて遷移させ、遷移後のステートに対応する行動を、次にとるべき行動として決定する。

【0042】ここで、行動決定機構部 52 は、所定のトリガ(trigger)があったことを検出すると、ステートを遷移させる。即ち、行動決定機構部 52 は、例えば、現在のステートに対応する行動を実行している時間が所定時間に達したときや、特定の状態認識情報を受信したとき、モデル記憶部 51 から供給される状態情報が示す感情や、本能、成長の状態の値が所定の閾値以下または以上になったとき等に、ステートを遷移させる。

【0043】なお、行動決定機構部 52 は、上述したように、センサ入力処理部 50 からの状態認識情報だけでなく、モデル記憶部 51 における感情モデルや、本能モデル、成長モデルの値等にも基づいて、行動モデルにおけるステートを遷移させることから、同一の状態認識情報が入力されても、感情モデルや、本能モデル、成長モデルの値（状態情報）によっては、ステートの遷移先は異なるものとなる。

【0044】その結果、行動決定機構部 52 は、例えば、状態情報が、「怒っていない」こと、および「お腹がすいていない」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「お手」という行動をとらせる行動指令情報を生成し、これを、姿勢遷移機構部 53 に送出する。

【0045】また、行動決定機構部 52 は、例えば、状態情報が、「怒っていない」こと、および「お腹がすいている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「手のひらをべろべろなめる」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部 53 に送出する。

【0046】また、行動決定機構部 52 は、例えば、状

態情報が、「怒っている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、状態情報が、「お腹がすいている」ことを表している、また、「お腹がすいていない」ことを表している、また、「ぶいと横を向く」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部 53 に送出する。

【0047】なお、行動決定機構部 52 には、モデル記憶部 51 から供給される状態情報が示す感情や、本能、成長の状態に基づいて、遷移先のステートに対応する行動のパラメータとしての、例えば、歩行の速度や、手足を動かす際の動きの大きさおよび速度などを決定させることができ、この場合、それらのパラメータを含む行動指令情報が、姿勢遷移機構部 53 に送出される。

【0048】また、行動決定機構部 52 では、上述したように、ロボットの頭部や手足等を動作させる行動指令情報の他、ロボットに発話を行わせる行動指令情報も生成される。ロボットに発話を行わせる行動指令情報は、音声合成部 55 に供給されるようになっている。この、音声合成部 55 に供給される行動指令情報には、音声合成部 55 に生成させる合成音に対応するテキスト等が含まれる。そして、音声合成部 55 は、行動決定部 52 から行動指令情報を受信すると、その行動指令情報に含まれるテキストに基づき、合成音を生成し、スピーカ 18 に供給して出力させる。これにより、スピーカ 18 からは、例えば、ロボットの鳴き声、さらには、「お腹がすいた」等のユーザへの各種の要求、「何？」等のユーザの呼びかけに対する応答その他の音声出力が行われる。ここで、音声合成部 55 には、モデル記憶部 51 から状態情報も供給されるようになっており、音声合成部 55 は、この状態情報が示す感情の状態に基づいて音質を制御した合成音を生成することが可能となっている。なお、音声合成部 55 では、感情の他、本能や成長の状態に基づいて音質を制御した合成音を生成することも可能である。

【0049】姿勢遷移機構部 53 は、行動決定機構部 52 から供給される行動指令情報に基づいて、ロボットの姿勢を、現在の姿勢から次の姿勢に遷移させるための姿勢遷移情報を生成し、これを制御機構部 54 に送出する。

【0050】ここで、現在の姿勢から次に遷移可能な姿勢は、例えば、胴体や手や足の形状、重さ、各部の結合状態のようなロボットの物理的形状と、関節が曲がる方向や角度のようなアクチュエータ 3A<sub>1</sub>乃至 5A<sub>1</sub>および 5A<sub>2</sub>の機構とによって決定される。

【0051】また、次の姿勢としては、現在の姿勢から直接遷移可能な姿勢と、直接には遷移できない姿勢とがある。例えば、4本足のロボットは、手足を大きく投げ出して寝転んでいる状態から、伏せた状態へ直接遷移することはできるが、立った状態へ直接遷移することはで

10

20

30

40

50

きず、一旦、手足を胴体近くに引き寄せて伏せた姿勢になり、それから立ち上がるという２段階の動作が必要である。また、安全に実行できない姿勢も存在する。例えば、４本足のロボットは、その４本足で立っている姿勢から、両前足を挙げてバンザイをしようとすると、簡単に転倒してしまう。

【0052】このため、姿勢遷移機構部５３は、直接遷移可能な姿勢をあらかじめ登録しておき、行動決定機構部５２から供給される行動指令情報が、直接遷移可能な姿勢を示す場合には、その行動指令情報を、そのまま姿勢遷移情報として、制御機構部５４に送出する。一方、行動指令情報が、直接遷移不可能な姿勢を示す場合には、姿勢遷移機構部５３は、遷移可能な他の姿勢に一旦遷移した後に、目的の姿勢まで遷移させるような姿勢遷移情報を生成し、制御機構部５４に送出する。これによりロボットが、遷移不可能な姿勢を無理に実行しようとする事態や、転倒するような事態を回避することができるようになっている。

【0053】制御機構部５４は、姿勢遷移機構部５３からの姿勢遷移情報にしたがって、アクチュエータ３Ａ<sub>１</sub>乃至５Ａ<sub>１</sub>および５Ａ<sub>２</sub>を駆動するための制御信号を生成し、これを、アクチュエータ３Ａ<sub>１</sub>乃至５Ａ<sub>１</sub>および５Ａ<sub>２</sub>に送出する。これにより、アクチュエータ３Ａ<sub>１</sub>乃至５Ａ<sub>１</sub>および５Ａ<sub>２</sub>は、制御信号にしたがって駆動し、ロボットは、自律的に行動を起こす。

【0054】次に、図４は、図３の音声認識部５０Ａの構成例を示している。

【0055】マイク１５からの音声信号は、ＡＤ(Analog Digital)変換部２１に供給される。ＡＤ変換部２１では、マイク１５からのアナログ信号である音声信号がサンプリング、量子化され、ディジタル信号である音声データにＡ／Ｄ変換される。この音声データは、特徴抽出部２２および音声区間検出部２７に供給される。

【0056】特徴抽出部２２は、そこに入力される音声データについて、適当なフレームごとに、例えば、ＭＦＣＣ(Mel Frequency Cepstrum Coefficient)分析を行い、その分析の結果得られるＭＦＣＣを、特徴パラメータ(特徴ベクトル)として、マッチング部２３に出力する。なお、特徴抽出部２２では、その他、例えば、線形予測係数、ケプストラム係数、線スペクトル対、所定の周波数帯域ごとのパワー(フィルタバンクの出力)等を、特徴パラメータとして抽出することが可能である。

【0057】マッチング部２３は、特徴抽出部２２からの特徴パラメータを用いて、音響モデル記憶部２４、辞書記憶部２５、および文法記憶部２６を必要に応じて参照しながら、マイク１５に入力された音声(入力音声)を、例えば、連続分布ＨＭＭ(Hidden Markov Model)法に基づいて音声認識する。

【0058】即ち、音響モデル記憶部２４は、音声認識する音声の言語における個々の音素や音節などの音響的

な特徴を表す音響モデルを記憶している。ここでは、連続分布ＨＭＭ法に基づいて音声認識を行うので、音響モデルとしては、ＨＭＭ(Hidden Markov Model)が用いられる。辞書記憶部２５は、認識対象の各単語について、その発音に関する情報(音韻情報)が記述された単語辞書を記憶している。文法記憶部２６は、辞書記憶部２５の単語辞書に登録されている各単語が、どのように連鎖する(つながる)かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法(CFG)や、統計的な単語連鎖確率(N-gram)などに基づく規則を用いることができる。

【0059】マッチング部２３は、辞書記憶部２５の単語辞書を参照することにより、音響モデル記憶部２４に記憶されている音響モデルを接続することで、単語の音響モデル(単語モデル)を構成する。さらに、マッチング部２３は、幾つかの単語モデルを、文法記憶部２６に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、連続分布ＨＭＭ法によって、マイク１５に入力された音声の認識する。即ち、マッチング部２３は、特徴抽出部２２が出力する時系列の特徴パラメータが観測されるスコア(尤度)が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列の音韻情報(読み)を、音声の認識結果として出力する。

【0060】より具体的には、マッチング部２３は、接続された単語モデルに対応する単語列について、各特徴パラメータの出現(出力)確率を累積し、その累積値をスコアとして、そのスコアを最も高くする単語列の音韻情報を、音声認識結果として出力する。

【0061】以上のようにして出力される、マイク１５に入力された音声の認識結果は、状態認識情報として、モデル記憶部５１および行動決定機構部５２に出力される。

【0062】なお、音声区間検出部２７は、ＡＤ変換部２１からの音声データについて、例えば、特徴抽出部２２がＭＦＣＣ分析を行うのと同様のフレームごとに、パワーを算出している。さらに、音声区間検出部２７は、各フレームのパワーを、所定の閾値と比較し、その閾値以上のパワーを有するフレームで構成される区間を、ユーザの音声が入力されている音声区間として検出する。そして、音声区間検出部２７は、検出した音声区間を、特徴抽出部２２とマッチング部２３に供給しており、特徴抽出部２２とマッチング部２３は、音声区間のみを対象に処理を行う。

【0063】次に、図５は、図３の音声合成部５５の構成例を示している。

【0064】テキスト解析部３１には、行動決定機構部５２が出力する、音声合成の対象とするテキストを含む行動指令情報が供給されるようになっており、テキスト解析部３１は、辞書記憶部３４や生成用文法記憶部３５

10

20

30

40

50



を参照しながら、その行動指令情報に含まれるテキストを解析する。

【0065】即ち、辞書記憶部34には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、生成用文法記憶部35には、辞書記憶部34の単語辞書に記述された単語について、単語連鎖に関する制約等の生成用文法規則が記憶されている。そして、テキスト解析部31は、この単語辞書および生成用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等のテキスト解析（言語解析）を行い、後段の規則合成部32で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセント、イントネーション、パワー等を制御するための韻律情報、各単語の発音を表す音韻情報などがある。

【0066】テキスト解析部31で得られた情報は、規則合成部32に供給され、規則合成部32は、音声情報記憶部36を参照しながら、テキスト解析部31に入力されたテキストに対応する合成音の音声データ（デジタルデータ）を生成する。

【0067】即ち、音声情報記憶部36には、例えば、CV(Consonant, Vowel)や、VCV、CVC、1ピッチ等の波形データの形で音素片データが、音声情報として記憶されており、規則合成部32は、テキスト解析部31からの情報に基づいて、必要な音素片データを接続し、さらに、音素片データの波形を加工することによって、ポーズ、アクセント、イントネーション等を適切に付加し、これにより、テキスト解析部31に入力されたテキストに対応する合成音の音声データ（合成音データ）を生成する。あるいは、また、音声情報記憶部36には、例えば、線形予測係数(LPC(Liner Prediction Coefficients))や、ケプストラム(cepstrum)係数等といった波形データを音響分析することにより得られる音声の特徴パラメータが、音声情報として記憶されており、規則合成部32は、テキスト解析部31からの情報に基づいて、必要な特徴パラメータを、音声合成用の合成フィルタのタップ係数として用い、さらに、その合成フィルタに与える駆動信号を出力する音源等を制御することによって、ポーズ、アクセント、イントネーション等を適切に付加し、これにより、テキスト解析部31に入力されたテキストに対応する合成音の音声データ（合成音データ）を生成する。

【0068】さらに、規則合成部32には、モデル記憶部51から状態情報が供給されるようになっており、規則合成部32は、その状態情報のうちの、例えば、感情モデルの値に基づいて、音声情報記憶部36に記憶された音声情報から、その音質を制御したものを生成し、あるいは、規則音声合成を制御する各種の合成制御パラメータを生成することによって、音質を制御した合成音データを生成する。

【0069】以上のようにして生成された合成音データは、スピーカ18に供給され、これにより、スピーカ18からは、テキスト解析部31に入力されたテキストに対応する合成音が、感情に応じて音質を制御して出力される。

【0070】なお、図3の行動決定機構部52では、上述したように、行動モデルに基づいて、次の行動が決定されるが、合成音として出力するテキストの内容は、ロボットの行動と対応付けておくことが可能である。

10 【0071】即ち、例えば、ロボットが、座った状態から、立った状態になる行動には、テキスト「よっこいしょ」などを対応付けておくことが可能である。この場合、ロボットが、座っている姿勢から、立つ姿勢に移行するときに、その姿勢の移行に同期して、合成音「よっこいしょ」を出力することが可能となる。

【0072】次に、図6は、図5の規則合成部32の構成例を示している。

20 【0073】韻律生成部41には、テキスト解析部31（図5）によるテキスト解析結果が供給され、韻律生成部41は、そのテキスト解析結果に基づいて、合成音の韻律を制御する韻律情報を生成する。韻律生成部41で生成された韻律情報は、波形生成部42に供給される。ここで、韻律制御部41では、合成音を構成する各音韻の継続時間長、合成音のピッチ周期の時間変化パターンを表す周期パターン信号、合成音のパワーの時間変化パターンを表すパワーパターン信号等が、韻律情報として生成される。

30 【0074】波形生成部42には、上述したように、韻律情報が供給される他、テキスト解析部31（図5）によるテキスト解析結果が供給される。さらに、波形生成部42には、パラメータ生成部43から合成制御パラメータが供給される。波形生成部42は、テキスト解析結果に含まれる音韻情報にしたがって、必要な変換音声情報を、変換音声情報記憶部45から読み出し、その変換音声情報を用いて規則音声合成を行うことにより、合成音を生成する。さらに、波形生成部42は、規則音声合成を行う際、韻律生成部41からの韻律情報と、パラメータ生成部43からの合成制御パラメータに基づいて、合成音データの波形を調整することにより、合成音の韻律と音質を制御する。そして、波形生成部42は、最終的に得られた合成音データを出力する。

【0075】パラメータ生成部43には、モデル記憶部51（図3）から状態情報が供給されるようになっており、パラメータ生成部43は、その状態情報のうちの感情モデルに基づいて、波形生成部42における規則音声合成を制御するための合成制御パラメータや、音声情報記憶部36（図5）に記憶された音声情報を変換する変換パラメータを生成する。

40 【0076】即ち、パラメータ生成部43は、例えば、感情モデルとしての「うれしさ」、「悲しさ」、「怒



り」、「楽しさ」、「興奮」、「眠い」、「心地よい」、「不快」等の感情の状態を表す値（以下、適宜、感情モデル値という）に、合成制御パラメータと変換パラメータを対応付けた変換テーブルを記憶しており、その変換テーブルにおいて、モデル記憶部 51 からの状態情報における感情モデルの値に対応付けられている合成制御パラメータと変換パラメータを出力する。

【0077】なお、パラメータ生成部 43 が記憶している変換テーブルは、ペットロボットの感情の状態を表す音質の合成音が得られるように、感情モデル値と、合成制御パラメータおよび変換パラメータとを対応付けて構成されている。感情モデル値と、合成制御パラメータおよび変換パラメータとを、どのように対応付けるかは、例えば、シミュレーションを行うことによって決定することができる。

【0078】さらに、ここでは、変換テーブルを用いて、感情モデル値から、合成制御パラメータおよび変換パラメータを得るようにしたが、その他、合成制御パラメータおよび変換パラメータは、例えば、次のようにして得ることも可能である。

【0079】即ち、例えば、ある感情 # n の感情モデル値を  $P_n$  と、ある合成制御パラメータまたは変換パラメータを  $Q_i$  と、所定の関数を  $f_{i,n}()$  と、それぞれ表すとき、合成制御パラメータまたは変換パラメータ  $Q_i$  は、式  $Q_i = \sum f_{i,n}(P_n)$  を計算することによって求めることが可能である。但し、 $\sum$  は、変数 n についてのサメーションを表す。

【0080】また、上述の場合には、「うれしさ」、「悲しさ」、「怒り」、「楽しさ」等のすべての感情モデル値を考慮した変換テーブルを用いるようにしたが、その他、例えば、次のような簡略化した変換テーブルを用いることも可能である。

【0081】即ち、感情の状態を、例えば、「定常」、「悲しさ」、「怒り」、「楽しさ」等のいずれかのみに分類し、各感情に、ユニークな番号としての感情番号を付しておく。即ち、例えば、「定常」、「悲しさ」、「怒り」、「楽しさ」に、それぞれ 0, 1, 2, 3 等の感情番号を、それぞれ付しておく。そして、このような感情番号と、合成制御パラメータおよび変換パラメータとを対応付けた変換テーブルを作成する。なお、このような変換テーブルを用いる場合には、感情モデル値から、感情の状態を、「うれしさ」、「悲しさ」、「怒り」、「楽しさ」のいずれかに分類する必要があるが、これは、次のようにして行うことが可能である。即ち、例えば、複数の感情モデル値のうち、最も大きい感情モデル値と、2 番目に大きい感情モデル値との差が、所定の閾値以上の場合は、最も大きい感情モデル値に対応する感情の状態に分類し、そうでない場合は、「定常」の状態に分類すればよい。

【0082】ここで、パラメータ生成部 43 において生

成される合成制御パラメータには、例えば、有声音や無声摩擦音、破裂音等の各音の音量バランスを調整するパラメータ、波形生成部 42 における音源としての、後述する駆動信号生成部 60（図 8）の出力信号の振幅ゆらぎの大きさを制御するパラメータ、音源の周波数を制御するパラメータ等の合成音の音質に影響するパラメータが含まれる。

【0083】また、パラメータ生成部 43 において生成される変換パラメータは、合成音を構成する波形データ 10 の特性を変更するように、音声情報記憶部 36（図 5）の音声情報を変換するためのものである。

【0084】パラメータ生成部 43 が生成する合成制御パラメータは、波形生成部 42 に供給され、変換パラメータは、データ変換部 44 に供給されるようになっている。データ変換部 44 は、音声情報記憶部 36 から音声情報を読み出し、変換パラメータにしたがって、音声情報を変換する。データ変換部 44 は、これにより、合成音を構成する波形データの特性を変更させる音声情報としての変換音声情報を得て、変換音声情報記憶部 45 に供給する。変換音声情報記憶部 45 は、データ変換部 44 から供給される変換音声情報を記憶する。この変換音声情報は、波形生成部 42 によって、必要に応じて読み出される。

【0085】次に、図 7 のフローチャートを参照して、図 6 の規則合成部 32 の処理について説明する。

【0086】図 5 のテキスト解析部 31 が出力するテキスト解析結果は、韻律生成部 41 と波形生成部 42 に供給される。また、図 5 のモデル記憶部 51 が出力する状態情報は、パラメータ生成部 43 に供給される。

【0087】韻律生成部 41 は、テキスト解析結果を受信すると、ステップ S1 において、テキスト解析結果に含まれる音韻情報が表す各音韻の継続時間長、周期パターン信号、パワーパターン信号等の韻律情報を生成し、波形生成部 42 に供給して、ステップ S2 に進む。

【0088】その後、ステップ S2 では、パラメータ生成部 43 は、感情反映モードかどうかを判定する。即ち、本実施の形態では、感情を反映した音質の合成音を出力する感情反映モードと、感情を反映しない音質の合成音を出力する非感情反映モードのうちのいずれかを設定することができるようになっており、ステップ S2 では、ロボットのモードが感情反映モードとなっているかどうか判定される。

【0089】ここで、ロボットには、感情反映モードと非感情反映モードを設けずに、常に、感情を反映した合成音を出力させるようにすることも可能である。

【0090】ステップ S2 において、感情反映モードでないと判定された場合、ステップ S3 および S4 をスキップして、ステップ S5 に進み、波形生成部 42 は、合成音を生成し、処理を終了する。

【0091】即ち、感情反映モードでない場合、パラメ

ータ生成部 43 は、特に処理を行わず、従って、合成制御パラメータおよび変換パラメータを生成しない。

【0092】その結果、波形生成部 42 は、音声情報記憶部 36 (図 5) に記憶された音声情報を、データ変換部 44 および変換音声情報記憶部 45 を介して読み出し、その音声情報と、デフォルトの合成制御パラメータを用い、韻律生成部 41 からの韻律情報に対応して韻律を制御しながら音声合成処理を行う。従って、波形生成部 42 では、デフォルトの音質を有する合成音データが生成される。

【0093】一方、ステップ S2 において、感情反映モードであると判定された場合、ステップ S3 に進み、パラメータ生成部 43 は、モデル記憶部 51 からの状態情報のうちの感情モデルに基づいて、合成制御パラメータおよび変換パラメータを生成する。そして、合成制御パラメータは、波形生成部 42 に供給され、変換パラメータは、データ変換部 44 に供給される。

【0094】その後、ステップ S4 に進み、データ変換部 44 が、パラメータ生成部 43 からの変換パラメータにしたがい、音声情報記憶部 36 (図 5) に記憶された音声情報を変換する。さらに、データ変換部 44 は、その変換の結果得られた変換音声情報を、変換音声情報記憶部 45 に供給して記憶させる。

【0095】そして、ステップ S5 に進み、波形生成部 42 は、合成音を生成し、処理を終了する。

$$s_n + \alpha_1 s_{n-1} + \alpha_2 s_{n-2} + \dots + \alpha_p s_{n-p} = e_n$$

で示す線形 1 次結合が成立すると仮定し、現在時刻 n のサンプル値  $s_n$  の予測値 (線形予測値)  $s_n'$  を、過去の

$$s_n' = -(\alpha_1 s_{n-1} + \alpha_2 s_{n-2} + \dots + \alpha_p s_{n-p})$$

によって線形予測したときに、実際のサンプル値  $s_n$  と線形予測値  $s_n'$  との間の自乗誤差を最小にする線形予測係数  $\alpha_p$  を求めるものである。

【0100】ここで、式 (1) において、 $\{e_n\}$  (・・

$$s_n = e_n - (\alpha_1 s_{n-1} + \alpha_2 s_{n-2} + \dots + \alpha_p s_{n-p})$$

で表すことができ、これを、Z 変換すると、次式が成立する。

$$S = E / (1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_p z^{-p})$$

但し、式 (4) において、S と E は、式 (3) における  $s_n$  と  $e_n$  の Z 変換を、それぞれ表す。

$$e_n = s_n - s_n'$$

で表すことができ、実際のサンプル値  $s_n$  と線形予測値  $s_n'$  との間の残差信号と呼ばれる。

【0104】従って、式 (4) から、線形予測係数  $\alpha_p$  を IIR (Infinite Impulse Response) フィルタのタ

ップ係数とするとともに、残差信号  $e_n$  を IIR フィル

【0096】即ち、この場合、波形生成部 42 は、変換音声情報記憶部 45 に記憶された音声情報のうちの必要なものを読み出し、その変換音声情報と、パラメータ生成部 43 から供給される合成制御パラメータを用い、韻律生成部 41 からの韻律情報に対応して韻律を制御しながら音声合成処理を行う。従って、波形生成部 42 では、ロボットの感情の状態に対応する音質を有する合成音データが生成される。

【0097】以上のように、感情モデル値に基づき、合成制御パラメータや変換パラメータを生成し、その合成制御パラメータや、変換パラメータによって音声情報を変換した変換音声情報を用いて音声合成を行うようにしたので、感情に応じて、例えば、周波数特性や音量バランス等といった音質が制御された、感情豊かな合成音を得ることができる。

【0098】次に、図 8 は、音声情報記憶部 36 (図 5) に記憶されている音声情報が、音声の特徴パラメータとしての、例えば線形予測係数 (LPC) である場合の、図 6 の波形生成部 42 の構成例を示している。

【0099】ここで、線形予測係数は、音声の波形データから求められた自己相関係数を用いた Yule-Walker の方程式を解く等の、いわゆる線形予測分析を行うことで得られるが、この線形予測分析は、現在時刻 n の音声信号 (のサンプル値)  $s_n$ 、およびこれに隣接する過去の P 個のサンプル値  $s_{n-1}$ ,  $s_{n-2}$ ,  $\dots$ ,  $s_{n-p}$  に、式

$$\dots (1)$$

P 個の標本値  $s_{n-1}$ ,  $s_{n-2}$ ,  $\dots$ ,  $s_{n-p}$  を用いて、式

$$\dots (2)$$

・,  $e_{n-1}$ ,  $e_n$ ,  $e_{n+1}$ ,  $\dots$ ) は、平均値が 0 で、分散が所定値  $\sigma^2$  の互いに無相関な確率変数である。

【0101】式 (1) から、サンプル値  $s_n$  は、式

$$\dots (3)$$

【0102】

$$\dots (4)$$

【0103】ここで、式 (1) および (2) から、 $e_n$  は、式

$$\dots (5)$$

タの駆動信号 (入力信号) とすることにより、音声信号  $s_n$  を求めることができる。

【0105】図 8 の波形生成部 42 は、式 (4) にしたがって音声信号を生成する音声合成を行うようになっている。

【0106】即ち、駆動信号生成部60は、駆動信号となる残差信号を生成して出力する。

【0107】ここで、駆動信号生成部60には、韻律情報、テキスト解析結果、および合成制御パラメータが供給されるようになっている。そして、駆動信号生成部60は、これらの韻律情報、テキスト解析結果、および合成制御パラメータにしたがい、周期（周波数）や振幅等を制御した周期的なインパルスと、ホワイトノイズのような信号とを重畳することにより、合成音に対して、対応する韻律、音韻、音質（声質）を与える駆動信号を生成する。なお、周期的なインパルスは、主として有声音の生成に寄与し、ホワイトノイズのような信号は、主として無声音の生成に寄与する。

【0108】図8において、1つの加算器61、P個の遅延回路（D）62<sub>1</sub>乃至62<sub>P</sub>、およびP個の乗算器63<sub>1</sub>乃至63<sub>P</sub>は、音声合成用の合成フィルタとしてのIIRフィルタを構成しており、駆動信号生成部60からの駆動信号を音源として、合成音データを生成する。

【0109】即ち、駆動信号生成部60が出力する残差信号（駆動信号）eは、加算器61を介して、遅延回路62<sub>1</sub>に供給され、遅延回路62<sub>1</sub>は、そこへの入力信号を、残差信号の1サンプル分だけ遅延して、後段の遅延回路62<sub>2</sub>に出力するとともに、演算器63<sub>1</sub>に出力する。乗算器63<sub>1</sub>は、遅延回路62<sub>1</sub>の出力と、そこにセットされた線形予測係数 $\alpha_1$ とを乗算し、その乗算値を、加算器61に出力する。

【0110】加算器61は、乗算器63<sub>1</sub>乃至63<sub>P</sub>の出力すべてと、残差信号eとを加算し、その加算結果を、遅延回路62<sub>1</sub>に供給する他、音声合成結果（合成音データ）として出力する。

$$F(\theta) = 1 / | 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_P z^{-P} |^2$$

$$z = e^{-j\theta}$$

【0117】周波数特性変換部73には、パラメータ生成部43（図6）が出力する変換パラメータが供給されるようになっている。そして、周波数特性変換部73は、フーリエ変換部72からのスペクトルを、変換パラメータにしたがって変換することにより、線形予測係数から得られる音声データ（波形データ）の周波数特性を変更する。

【0118】ここで、図9の実施の形態では、周波数特性変換部73は、伸縮処理部73Aとイコライザ73Bとから構成されている。

【0119】伸縮処理部73は、フーリエ変換部72から供給されるスペクトルF(θ)を、周波数軸方向に伸縮させる。即ち、伸縮処理部73Aは、伸縮パラメータをΔと表すと、式(6)を、そのθをΔθに替えて演算し、周波数軸方向に伸縮を行ったスペクトルF(Δθ)を求める。

【0120】この場合、伸縮パラメータΔが、変換パラ

【0111】なお、係数供給部64は、変換音声情報記憶部45から、テキスト解析結果に含まれる音韻等に応じて、必要な変換音声情報としての線形予測係数 $\alpha_1, \alpha_2, \dots, \alpha_P$ を読み出し、それぞれを、乗算器63<sub>1</sub>乃至63<sub>P</sub>にセットするようになっている。

【0112】次に、図9は、音声情報記憶部36（図5）に記憶されている音声情報が、音声の特徴パラメータとしての、例えば、線形予測係数（LPC）である場合の、図6のデータ変換部44の構成例を示している。

【0113】音声情報記憶部36に記憶された音声情報としての線形予測係数は、合成フィルタ71に供給される。合成フィルタ71は、図8における1つの加算器61、P個の遅延回路（D）62<sub>1</sub>乃至62<sub>P</sub>、およびP個の乗算器63<sub>1</sub>乃至63<sub>P</sub>でなる合成フィルタと同様のIIRフィルタであり、線形予測係数をタップ係数とするとともに、インパルスを駆動信号としてフィルタリングを行うことで、線形予測係数を音声データ（時領域の波形データ）に変換する。この音声データは、フーリエ変換部72に供給される。

【0114】フーリエ変換部72は、合成フィルタ71からの音声データをフーリエ変換することにより、周波数領域の信号、即ち、スペクトルを求め、周波数特性変換部73に供給する。

【0115】従って、合成フィルタ71およびフーリエ変換部72では、線形予測係数 $\alpha_1, \alpha_2, \dots, \alpha_P$ がスペクトルF(θ)に変換されるが、この線形予測係数 $\alpha_1, \alpha_2, \dots, \alpha_P$ からスペクトルへF(θ)の変換は、その他、例えば、次式にしたがい、θを0からπまでに変化させることによっても行うことができる。

30 【0116】

・・・(6)

メータとなる。なお、伸縮パラメータΔは、例えば、0.5乃至2.0の範囲内の値とすることができる。

【0121】イコライザ73Bは、フーリエ変換部72から供給されるスペクトルF(θ)に、イコライジング処理を施すことにより、その高域を強調または抑圧する。即ち、イコライザ73Bは、スペクトルF(θ)に対して、例えば、図10(A)に示すような特性の高域強調フィルタ、または図10(B)に示すような特性の高域抑圧フィルタをかけ、その周波数特性を変更したスペクトルを求める。

【0122】ここで、図10において、gはゲインを、f<sub>1</sub>は遮断周波数を、f<sub>2</sub>は減衰幅を、f<sub>3</sub>は音声データ（合成フィルタ71が出力する音声データ）のサンプリング周波数を、それぞれ表すが、このうちのゲインg、遮断周波数f<sub>1</sub>、および減衰幅f<sub>2</sub>が、変換パラメータとなる。

【0123】なお、一般に、図10(A)の高域強調フ

ィルタをかけた場合には、合成音の音質は、固い印象のものとなり、図 10 (B) の高域抑圧フィルタをかけた場合には、合成音の音質は、柔らかい印象のものとなる。

【0124】また、周波数特性変換部 73 では、その他、例えば、 $n$  次平均フィルタをかけたり、ケプストラム係数を求めてリフタ(lifter)をかける等して、スペクトルを平滑化することも可能である。

【0125】周波数特性変換部 73 において周波数特性の変換されたスペクトルは、逆フーリエ変換部 74 に供給される。逆フーリエ変換部 74 は、周波数特性変換部 73 からのスペクトルを逆フーリエ変換することにより、時領域の信号、即ち、音声データ(波形データ)を求め、LPC 分析部 75 に供給する。

【0126】LPC 分析部 75 は、逆フーリエ変換部 74 からの音声データを線形予測分析することにより、線形予測係数を求め、この線形予測係数を、変換音声情報として、変換音声情報記憶部 45 (図 6) に供給して記憶させる。

【0127】なお、ここでは、音声の特徴パラメータとして、線形予測係数を採用したが、その他、ケプストラム係数や、線スペクトル対等を採用することも可能である。

【0128】次に、図 11 は、音声情報記憶部 36 (図 5) に記憶されている音声情報が、音声データ(波形データ)としての、例えば音素片データである場合の、図 6 の波形生成部 42 の構成例を示している。

【0129】接続制御部 81 には、韻律情報、合成制御パラメータ、およびテキスト解析結果が供給されるようになっている。接続制御部 81 は、これらの韻律情報、合成制御パラメータ、およびテキスト解析結果にしたがい、合成音を生成するのに接続すべき音素片データや、その波形の調整方法を決定し、波形接続部 82 を制御する。

【0130】波形接続部 82 は、接続制御部 81 の制御にしたがい、変換音声情報記憶部 45 から、変換音声情報としての、必要な音素片データを読み出し、さらに、同じく接続制御部 81 の制御にしたがい、読み出した音素片データの波形を調整して接続する。これにより、波形接続部 82 は、韻律情報、合成制御パラメータ、テキスト解析結果それぞれに対応する韻律、音質、音韻の合成音データを生成して出力する。

【0131】次に、図 12 は、音声情報記憶部 36 (図 5) に記憶されている音声情報が、音声データ(波形データ)である場合の、図 6 のデータ変換部 44 の構成例を示している。なお、図中、図 9 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、図 12 のデータ変換部 44 は、合成フィルタ 71 および LPC 分析部 75 が設けられていない他は、図 9 における場合と同様に

構成されている。

【0132】従って、図 12 のデータ変換部 44 では、フーリエ変換部 72 において、音声情報記憶部 36 (図 5) に記憶された音声情報としての音声データがフーリエ変換され、その結果得られるスペクトルが、周波数特性変換部 73 に供給される。周波数特性変換部 73 は、フーリエ変換部 72 からのスペクトルに対して、変更パラメータにしたがった周波数特性変換処理を施し、逆フーリエ変換部 74 に出力する。逆フーリエ変換部 74 は、周波数特性変換部 73 からのスペクトルを逆フーリエ変換することにより、音声データとし、この音声データを、変換音声情報として、変換音声情報記憶部 45 (図 6) に供給して記憶させる。

【0133】以上、本発明を、エンターテイメント用のロボット(疑似ペットとしてのロボット)に適用した場合について説明したが、本発明は、これに限らず、例えば、音声合成装置を搭載した各種のシステムに広く適用することが可能である。また、本発明は、現実世界のロボットだけでなく、例えば、液晶ディスプレイ等の表示装置に表示される仮想的なロボットにも適用可能である。

【0134】なお、本実施の形態においては、上述した一連の処理を、CPU 10A にプログラムを実行させることにより行うようにしたが、一連の処理は、それ専用のハードウェアによって行うことも可能である。

【0135】ここで、プログラムは、あらかじめメモリ 10B (図 2) に記憶させておく他、フロッピー(登録商標)ディスク、CD-ROM (Compact Disc Read Only Memory)、MO (Magneto-optical) ディスク、DVD (Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体に、一時的あるいは永続的に格納(記録)しておくことができる。そして、このようなリムーバブル記録媒体を、いわゆるパッケージソフトウェアとして提供し、ロボット(メモリ 10B) にインストールするようにすることができる。

【0136】また、プログラムは、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、無線で転送したり、LAN (Local Area Network)、インターネットといったネットワークを介して、有線で転送し、メモリ 10B にインストールすることができる。

【0137】この場合、プログラムがバージョンアップされたとき等に、そのバージョンアップされたプログラムを、メモリ 10B に、容易にインストールすることができる。

【0138】なお、本明細書において、CPU 10A に各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理(例えば、並列処理あるいはオブジェクトによる処理)も含むものである。

【0139】また、プログラムは、1のCPUにより処理されるものであっても良いし、複数のCPUによって分散処理されるものであっても良い。

【0140】次に、図5の音声合成装置55は、専用のハードウェアにより実現することもできるし、ソフトウェアにより実現することもできる。音声合成装置55をソフトウェアによって実現する場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0141】そこで、図13は、音声合成装置55を実現するためのプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0142】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0143】あるいはまた、プログラムは、フロッピーディスク、CD-ROM、MOディスク、DVD、磁気ディスク、半導体メモリなどのリムーバブル記録媒体111に、一時的あるいは永続的に格納（記録）しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。

【0144】なお、プログラムは、上述したようなリムーバブル記録媒体111からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部108で受信し、内蔵するハードディスク105にインストールすることができる。

【0145】コンピュータは、CPU(Central Processing Unit)102を内蔵している。CPU102には、バス101を介して、入出力インタフェース110が接続されており、CPU102は、入出力インタフェース110を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部107が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM(Random Access Memory)104にロードして実行する。これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に応

じて、例えば、入出力インタフェース110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0146】なお、本実施の形態では、感情の状態に基づいて、合成音の音質を変えるようにしたが、その他、例えば、感情の状態に基づいて、合成音の韻律も変えるようにすることが可能である。合成音の韻律は、例えば、合成音のピッチ周期の時間変化パターン（周期パターン）や、合成音のパワーの時間変化パターン（パワーパターン）等を、感情モデルに基づいて制御することで変えることが可能である。

【0147】また、本実施の形態においては、テキスト（漢字仮名交じりのテキストを含む）から合成音を生成するようにしたが、その他、発音記号等から合成音を生成するようにすることも可能である。

【0148】

【発明の効果】以上の如く、本発明の音声合成装置および音声合成方法、並びにプログラムによれば、所定の情報のうち、合成音の音質に影響する音質影響情報が、外部から供給される、感情の状態を表す状態情報に基づいて生成され、その音質影響情報を用いて、音質を制御した合成音が生成される。従って、感情の状態に応じて音質を変えた合成音を生成することにより、感情豊かな合成音を得ることが可能となる。

【図面の簡単な説明】

【図1】本発明を適用したロボットの一実施の形態の外観構成例を示す斜視図である。

【図2】ロボットの内部構成例を示すブロック図である。

【図3】コントローラ10の機能的構成例を示すブロック図である。

【図4】音声認識部50Aの構成例を示すブロック図である。

【図5】音声合成部55の構成例を示すブロック図である。

【図6】規則合成部32の構成例を示すブロック図である。

【図7】規則合成部32の処理を説明するフローチャートである。

【図8】波形生成部42の第1の構成例を示すブロック図である。

【図9】データ変換部44の第1の構成例を示すブロック図である。

【図10】イコライザ73Bの処理を説明するための図である。

【図11】波形生成部42の第2の構成例を示すブロック図である。

【図12】データ変換部44の第2の構成例を示すブロ

23

ック図である。

【図 13】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

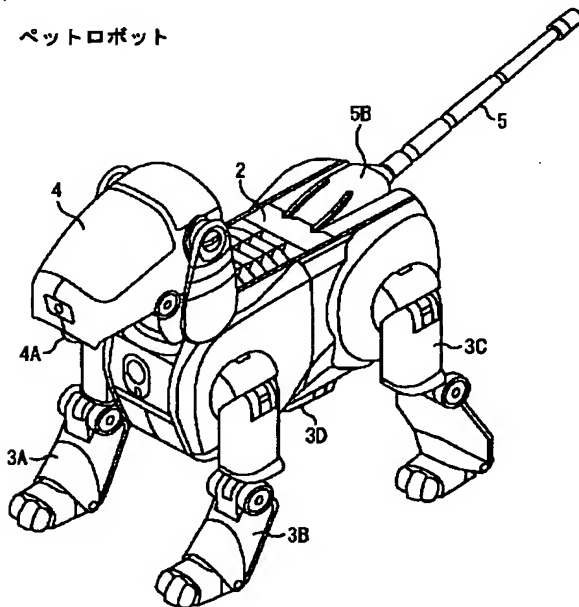
【符号の説明】

1 頭部ユニット, 4A 下顎部, 10 コントローラ, 10A CPU, 10B メモリ, 15 マイク, 16 CCDカメラ, 17 タッチセンサ, 18 スピーカ, 21 AD変換部, 22 特徴抽出部, 23 マッチング部, 24 音響モデル記憶部, 25 辞書記憶部, 26 文法記憶部, 27 音声区間検出部, 31 テキスト解析部, 32 規則合成部, 34 辞書記憶部, 35 生成用文法記憶部, 36 音声情報記憶部, 41 韻律生成部, 42 波形生成部, 43 パラメータ生成部, 44 データ変換部, 45 変換音声情報記憶

【図 1】

図1

ペットロボット



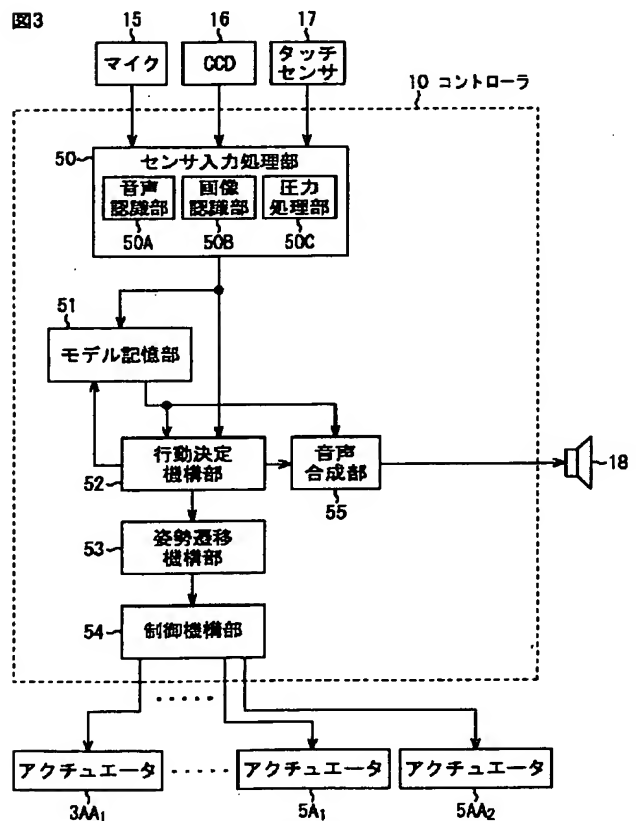
ペットロボットの外觀構成

24

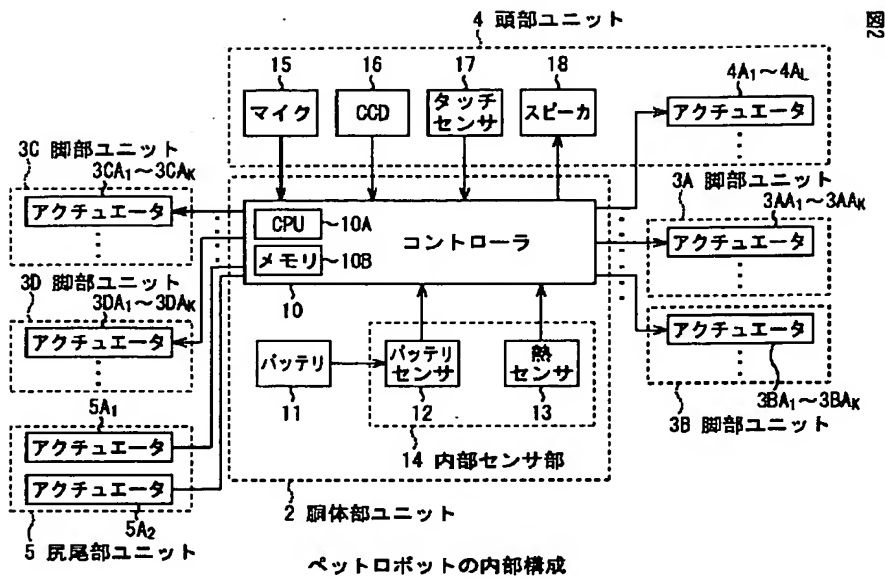
部, 50 センサ入力処理部, 50A 音声認識部, 50B 画像認識部, 50C 圧力処理部, 51 モデル記憶部, 52 行動決定機構部, 53 姿勢遷移機構部, 54 制御機構部, 55 音声合成部, 60 駆動信号生成部, 61 加算器, 62,乃至62, 遅延回路, 63,乃至63, 乗算器, 64 係数供給部, 71 合成フィルタ, 72 フーリエ変換部, 73 周波数特性変換部, 73A 伸縮処理部, 73B イコライザ, 74 逆フーリエ変換部, 75 LPC分析部, 81 接続制御部, 82 波形接続部, 101 バス, 102 CPU, 103 ROM, 104 RAM, 105 ハードディスク, 106 出力部, 107 入力部, 108 通信部, 109 ドライブ, 110 入出力インタフェース, 111 リムーバブル記録媒体

【図 3】

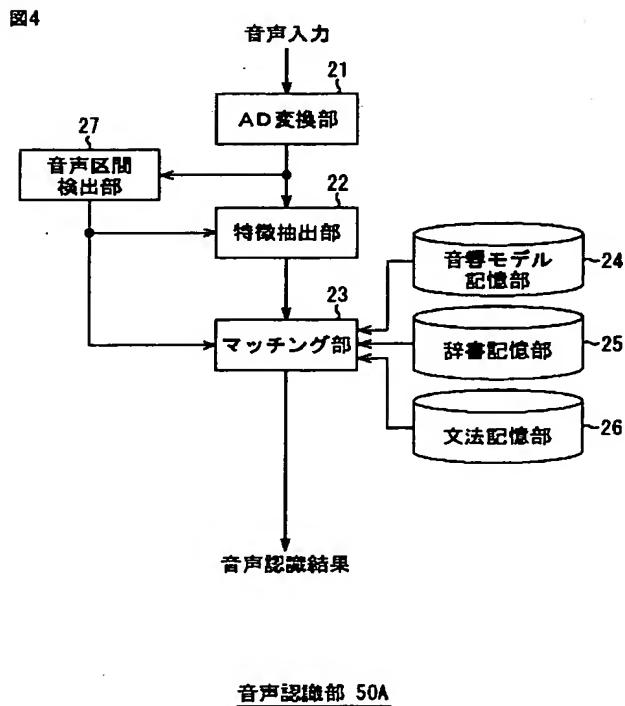
図3



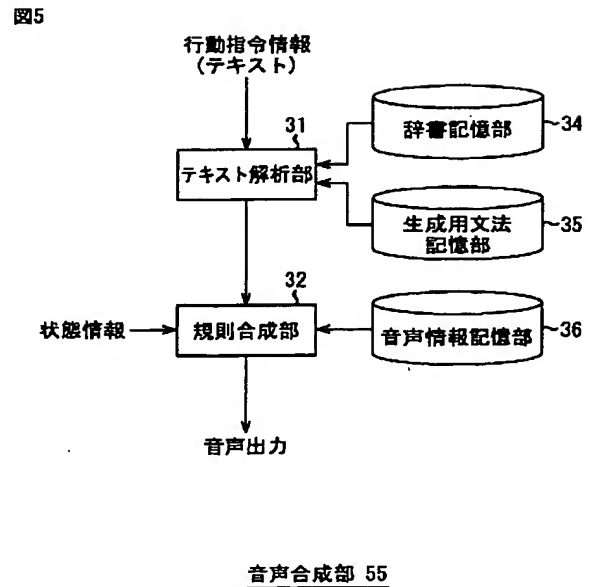
【図 2】



【図 4】

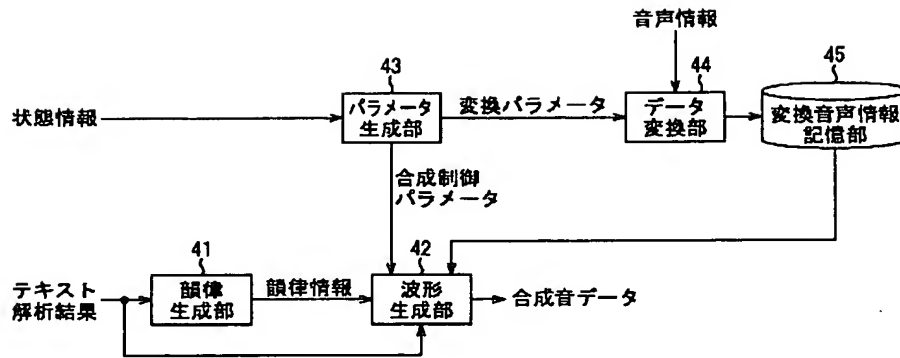


【図 5】





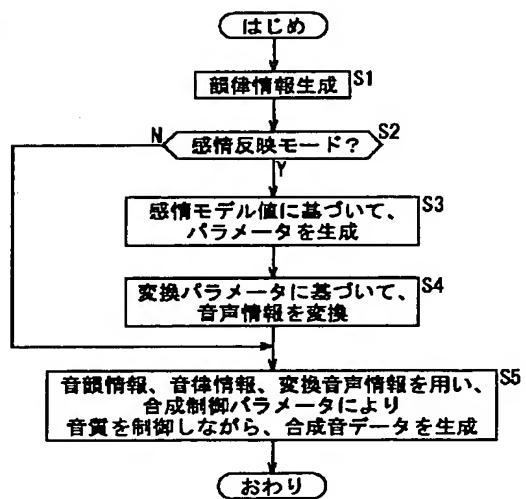
【図 6】



規則合成部 32

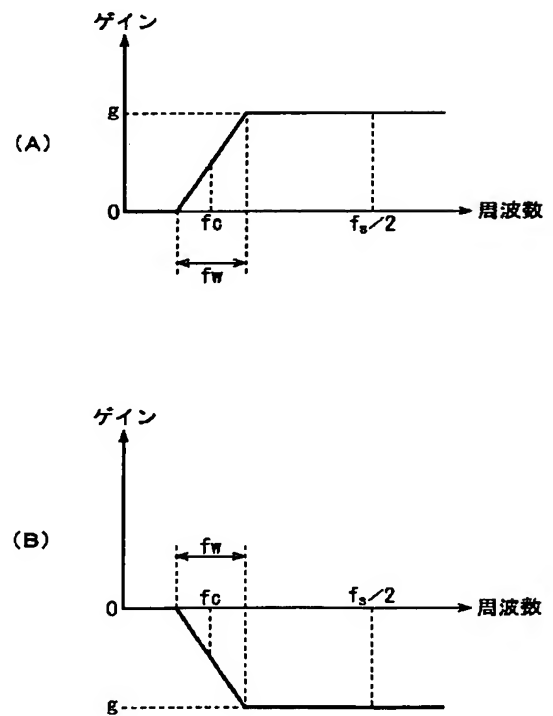
【図 7】

図7



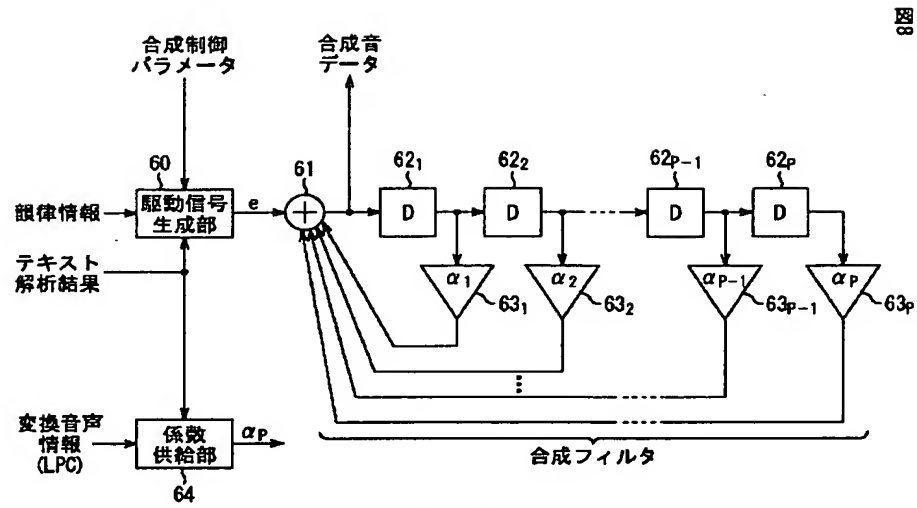
【図 10】

図10



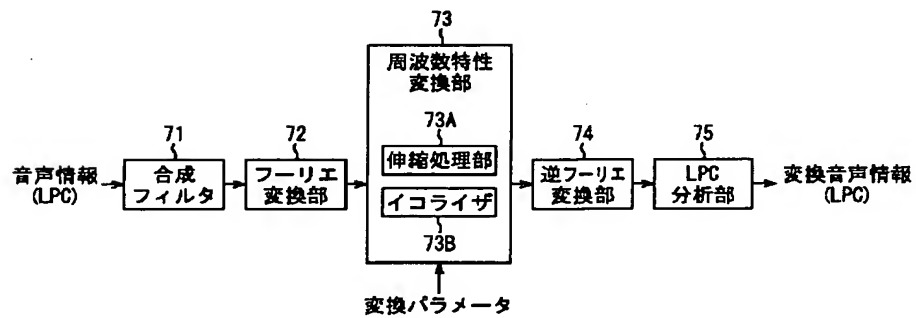
イコライザの処理

【図 8】



波形生成部 42

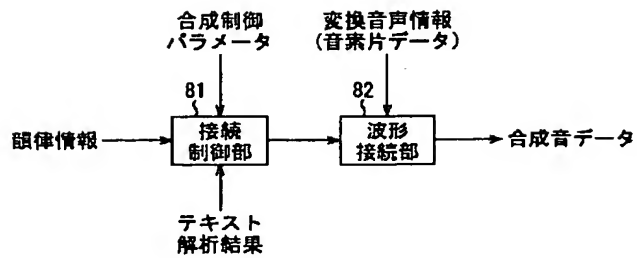
【図 9】



データ変換部 44

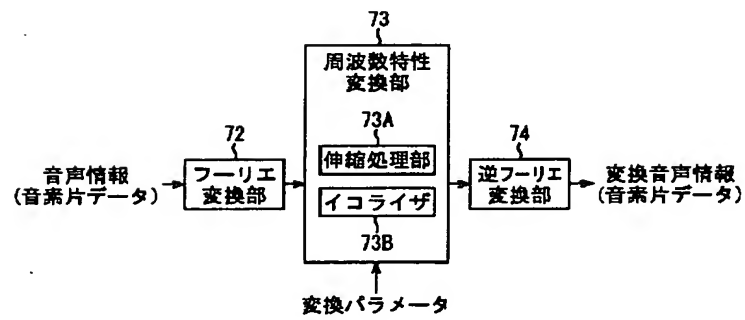
【図 11】

図 11

波形生成部 42

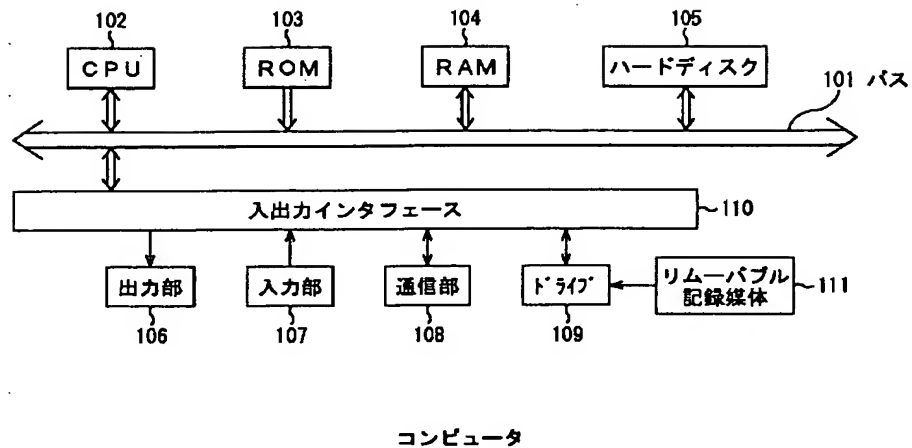
【図 12】

図 12

データ変換部 44

【図 13】

図 13



フロントページの続き

(72)発明者 浅野 康治  
 東京都品川区北品川 6 丁目 7 番 35 号 ソニ  
 ー株式会社内  
 (72)発明者 狩谷 真一  
 東京都品川区北品川 6 丁目 7 番 35 号 ソニ  
 ー株式会社内

(72)発明者 藤田 八重子  
 東京都品川区北品川 6 丁目 7 番 35 号 ソニ  
 ー株式会社内

F ターム(参考) 2C150 CA01 CA02 CA04 DA05 DA24  
 DA25 DA26 DA27 DA28 DF03  
 DF04 DF33 ED42 ED52 EF03  
 EF07 EF09 EF13 EF16 EF23  
 EF29 EF34 EF36  
 5D045 AA08 AA09 AB11